OFFICE OF NAVAL RESEARCH

FINAL REPORT FOR 1988–1990

for

CONTRACT N00014-88-K-0265 P00001

1 MAY 1988 – 30 SEPTEMBER 1990

Robust and Nonparametric Time Series, And Problems in Spatial Statistics

Principal Investigator: **R. Douglas Martin**
Co-Principal Investigator: **Adrian E. Raftery**

Department of Statistics, GN-22
University of Washington
Seattle, Washington 98195 USA

DTIC
ELECTE
AUG16 1993
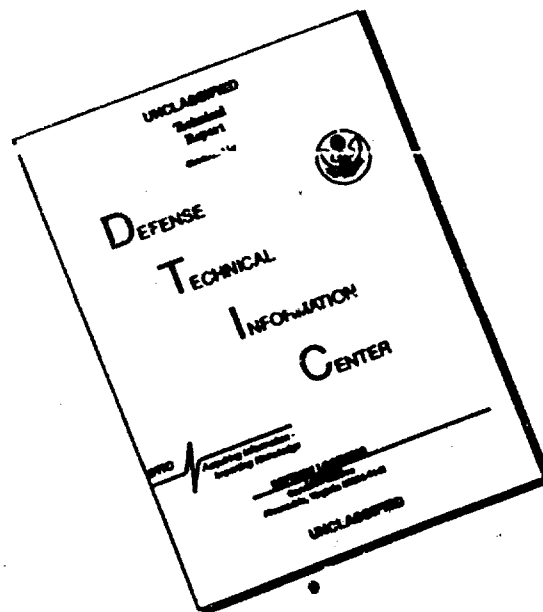S E D

DISCLAIMER NOTICE

THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

CONTENTS

Accesion For		
NTIS CRA&I		☑
DTIC TAB		☑
Unannounced		☐
Justification		
By		
Distribution /		
Availability Codes		
Dist	Avail and / or Special	
A-1		

DTIC QUALITY INSPECTED 3

1. INTRODUCTION

This final report describes the research results obtained under Office of Naval Research Contract N00014-88-K-0265 with the University of Washington, entitled *Robust and Non-parametric Time Series, and Problems in Spatial Statistics*. The Principal Investigators were R. Douglas Martin and Adrian E. Raftery, Department of Statistics, University of Washington. The research carried out represents a very broad and basic program dealing with robust, non-linear, non-Gaussian and long-memory time series, and with problems in spatial analysis. One major development during this contract period has been the use, for the first time, of clustering methods to delineate the boundaries of objects in images, and its successful application to the detection of ice floes.

This report is organized in the following way. Section 2 lists the major topic areas and sub-problem areas in which research results were obtained, along with the names of the various researchers who contributed to the effort. We have been particularly fortunate to have the participation of several truly outstanding researchers from other institutions on a regular basis at a level of funding of 1-2 months per year. Chief among these are Dennis Cox (University of Illinois), Victor Yohai (University of Buenos Aires) and Ruben Zamar (University of British Columbia). Also, Keith Knight (University of Toronto) and Ricardo Maronna (University of La Plata) participated, and the program included major collaboration with Jeff Banfield (University of Montana), who was supported by ONR's Polar Science Program.

Secction 3 lists invited papers and invited talks, which include two Royal Statistical Discussion papers and several invited discussions of other discussion papers. The contract resulted in 16 papers published or accepted for publication (including the two major Royal Statistical Society Discussion papers). An additional ten papers have been or soon will be submitted for publication.

Section 4 contains readable abstract-type descriptions of the research results obtained on each of the sub-problems in the various major areas. Reprints and Technical Reports containing the details will be provided under separate cover, if so desired by ONR.

Section 5 provides a complete listing of published papers, papers under review and technical reports, and finally Section 6 lists the graduate students who were supported by the contract, with dates of completion of Ph.D. indicated.

2. BASIC AREAS OF RESEARCH AND ASSOCIATED INVESTIGATIONS

CLUSTERING METHODS FOR IMAGE ANALYSIS

Ice floe identification in satellite images (Raftery)

Model-based Gaussian and non-Gaussian clustering (Raftery)

NON-GAUSSIAN TIME SERIES

Prediction, ergodicity and recurrence in general non-Gaussian state-space models (Raftery)

Time series of continuous proportions (Raftery)

Mixture transition distribution models (Le, Martin and Raftery)

Multivariate exponential and MPH distributions (Raftery)

ROBUSTNESS IN TIME SERIES

Leave-k-out diagnostics (Bruce and Martin)

Approximate non-Gaussian conditional mean filters (Fraiman and Martin)

Robust fixed lag smoothing (Le and Martin)

M-estimates for nearly nonstationary autoregression (Cox)

Robust model comparison for ARIMA models (Le, Raftery, Martin)

BIAS ROBUSTNESS AND ROBUSTNESS IN GENERAL MODELS

Bias robust estimates in regression (Martin, Yohai, Zamar, Maronna)

Bias robust M-estimates of scale for positive random variables (Martin and Zamar)

Estimation of score functions (Cox and Martin)

Min-max variance of M-estimates (Zamar)

LONG-MEMORY, INFINITE VARIANCE AND NEARLY NONSTATIONARY TIME SERIES MODELS

Long memory time series and spatial modeling (Raftery)

Infinite variance time series modeling (Knight)

Nearly nonstationary time series models (Cox)

BAYESIAN ANALYSIS OF DYNAMIC MODELS

Dynamic Bayesian analysis of software reliability models (Raftery)

Bayesian estimation of unknown population sizes (Raftery)

Bayesian nonparametric regression (Cox)

3. INVITED PAPERS, DISCUSSIONS, AND TALKS

Martin, R.D. (1989). "Leave-k-out Diagnostics for Time Series." (with A. Bruce) Royal Statistical Society Discussion Paper London.

Raftery, A.E. (1989). "A unified approach to N estimation." Institute of Mathematical Statistics Eastern Regional Meeting, Lexington, Kentucky.

Raftery, A.E. (1989). "Binomial N and bowhead whales: A unified approach to N estimation." Carnegie-Mellon University, Department of Statistics.

Raftery, A.E. (1989). "Forecasting with non-Gaussian time series" The Institute of Management Sciences/Operations Research Society of America, Annual Meeting, Vancouver, B.C. (organized session).

Raftery, A.E. (1990) Institute for Mathematics and its Applications (Minnesota): Special Summer Session on Time Series.

Martin, R.D. (1989). "Robustness Challenges in Time Series" Institute for Mathematics and its Applications (Minnesota): Special Summer Session on Robustness and Diagnostics. (This is one of a relatively small number of talks for the four week session).

Martin, R.D. (1989) "A Robust Strategy for Fitting ARIMA Models with Outliers and Level Shifts" (with A. Bruce and V. Yohai). American Statistical Association Special Session on Outliers in Time Series. Washington, D.C..

Martin, R.D. (1989). "Robustness Concepts for Time Series" ECAS One-Week Course on Robustness Germany, October, 1989. (Other key participants will include Hampel, Ronchetti, Rousseeuw and Stahel).

Martin, R.D. (1989). "Leave-k-out diagnostics for time series" (with A.G. Bruce). Royal Statistical Society, London.
[Note: This is a Royal Statistical Society Discussion Paper]

Martin, R.D. (1989). "Robustness problems in time series." IMA Summer workshop on Robustness and diagnostics, Minneapolis, Minn.

Martin, R.D. (1989). "Bias robust and high-breakdown point estimates of autoregressions." (with V.J. Yohai) Business and Economics Invited Session on Time Series, Annual ASA Meeting, Washington, D.C..

Martin, R.D. (1989). "Robust methods for time series." European Community Special Course on Robustness in Statistics, Schloss Reisensberg, Germany.

Raftery, A.E. (1989). "Forecasting with non-Gaussian time series." The Institute of Management Sciences/Operations Research Society of America Annual Meeting, Vancouver, B.C.. (organized session)

4. DESCRIPTION OF RESEARCH RESULTS

Ice Floe Identification in Satellite Imagery Using Mathematical Morphology and Clustering About Principal Curves
Banfield, J.D. and Raftery, A.E. (1989)

Identification of ice floes and their outlines in satellite images is important for understanding physical processes in the polar regions, for transportation in ice-covered seas and for the design of offshore structures intended to survive in the presence of ice. At present this is done manually, a long and tedious process which precludes full use of the great volume of relevant images now available.

We describe an automatic and accurate method for identifying ice floes and their outlines. Floe outlines are modeled as closed principal curves (Hastie and Stuetzle, 1989), a flexible class of smooth non-parametric curves. We propose a robust method of estimating closed principal curves which reduces both bias and variance. Initial estimates of floe outlines come from the erosion-propagation (EP) algorithm, which combines erosion from mathematical morphology with local propagation of information about floe edges.

The edge pixels from the EP algorithm are grouped into floe outlines using a new clustering algorithm. This extends existing clustering methods by allowing groups to be centered about principal curves rather than points or lines. This may open the way to efficient feature extraction using cluster analysis in images more generally. The method is implemented in an object-oriented programming environment for which it is well suited, and is quite computationally efficient.

Model-based Gaussian and non-Gaussian Clustering
Banfield, J.D. and Raftery, A.E. (1989)

The classification maximum likelihood approach is sufficiently general to encompass many current clustering algorithms, including those based on the sum of squares criterion and on the criterion of Friedman and Rubin (1967). However, as currently implemented, it does not allow the specification of which feature (orientation, size and shape) are to be common to all clusters and which may differ between clusters. Also, it is restricted to Gaussian distributions and it does not allow for noise.

We propose ways of overcoming these limitations. A reparameterization of the covariance matrix allows us to specify that some features, but not all, be the same for all clusters. A practical framework for non-Gaussian clustering is outlined, and a means of incorporating noise in the form of a Poisson process is described. An approximate Bayesian method for choosing the number of clusters is given.

The performance of the proposed methods is studied by simulation, with encouraging results. The methods are applied to the analysis of a data set arising in the study of diabetes, and the results seem better than those of previous analyses.

Leave-k-out Diagnostics for Time Series
Bruce, A.G., and Martin, R.D. (1989).

We propose diagnostics for autoregressive integrated moving average (ARIMA) model fitting for time series formed by deleting observations from the data and measuring the change in the estimates of the parameters. The use of leave-one-out diagnostics is a well-established tool in regression analysis. We demonstrate the efficacy of observation-deletion-based diagnostics for ARIMA models, addressing issues special to the time series setting. It is shown that the dependency aspect of time series data gives rise to a 'smearing' effect, which confounds the diagnostics for the coefficients. It is also shown that the diagnostics based on the innovations variance are much clearer and more sensitive than those for the coefficients. A 'leave-k-out' diagnostics approach is proposed to deal with patches of outliers, and problems cased by 'masking' are handled by use of iterative deletion. An overall strategy for ARIMA model fitting is given and applied to two data sets.

Gaussian Likelihood Estimation for Nearly Nonstationary AR(1) Processes
Cox, D. (1987)

An asymptotic analysis is presented for estimation in the three parameter first order autoregressive model, where the parameters are the mean, autoregressive coefficient, and variance of the shocks. the nearly nonstationary asymptotic model is considered wherein the autoregressive coefficient tends to 1 as sample size tends to infinity. Three different estimators are considered: the exact Gaussian maximum likelihood estimator, the conditional maximum likelihood or least squares estimator, and some "naive" estimators. It is shown that the estimators converge in distribution to analogous estimators for a continuous time Ornstein-Uhlenbeck process. Simulation results show that the MLE has smaller asymptotic mean squared error than the other two, and that the conditional maximum likelihood estimator gives a very poor estimator of the process mean.

An Analysis of Bayesian Inference for Nonparametric Regression
Cox, D. (1988)

The observation model $y_i = \beta(i/n) + \varepsilon_i$, $1 \leq i \leq n$, is considered, where the ε's are i.i.d. mean zero and variance σ^2 and β is an unknown smooth function. A Gaussian prior distribution is specified by assuming β is the solution of a high order stochastic differential equation. The estimation error $\delta = \beta - \hat{\beta}$ is analyzed, where $\hat{\beta}$ is the posterior expectation of β. Asymptotic posterior and sampling distributional approximations are given for $\| \delta \|^2$ when $\| \cdot \|$ is one of a family of norms natural to the problem. It is shown that the frequentist coverage probability of a variety of $(1-\alpha)$ posterior probability regions tends to be larger than $1-\alpha$, but will be infinitely often less than any $\varepsilon > 0$ as $n \to \infty$ with prior probability 1. A related continuous time signal estimation problem is also studied.

M-Estimates for Nearly Nonstationary Autoregression
Cox, D. and Llatas, E. (1988).

The nearly nonstationary first order autoregression is a sequence of processes where the autoregressive coefficient tends to 1 as $n \to \infty$. M-estimates of the autoregressive coefficient

are considered. The process is allowed to be nongaussian, but a 2+δ moment condition is assumed. the limiting distribution is not the usual normal limit but is characterized as a ratio of two stochastic integrals. the asymptotically most efficient M-estimate is not given by maximum likelihood. However, it is shown that the loss of efficiency in using maximum likelihood is no worse than about 20%, whereas the usual least squares estimator can have arbitrarily low efficiency.

Estimation of score functions: Bikernel estimates and cross-validation
Cox, D. and Martin, R.D. (1988)

The score function is the negative logarithmic derivative of the density: $\psi = -f'/f$. Estimates of the score function can be used for data analysis (assessing tail behavior, unimodality) as well as for adaptive estimates of location. Estimators of ψ based on kernel density estimators are proposed and analyzed. Bandwidth estimation is especially difficult. Several bandwidth estimators are proposed and investigated via Monte Carlo.

Approximate Non-Gaussian Conditional Mean Filters
Fraiman, R., and Martin, R.D. (1989)

Ergodicity and Recurrence for Exponential Family State Space Models
Grunwald, G.K., Raftery, A.E., and Guttorp, P. (1989)

We give two results concerning the properties of state-space models with exponential family observation distribution and conjugate state distribution. The first result gives a simple and general interpretation of the parameters of the predictive state distribution in terms of the observation forecast distribution. The second result shows how the first one can be used to check the long-term model properties of ergodicity and recurrence for a certain class of non-Gaussian observation distribution. In particular, these results apply to models with Poisson, binomial or multinomial observation distributions.

Time Series of Continuous Proportions
Grunwald, G.K., Raftery, A.E., and Guttorp, P. (1989)

A vector of continuous proportions consists of the proportions of some total accounted for by its constituent components. An example is the proportions of Social Security tax revenues from each of personal tax, corporate tax and social tax. We consider the situation where time series data are available and where interest focuses on the proportions rather than the actual amounts. Reasons for analyzing such time series include estimation of the underlying trend, estimation of the effect of covariates and interventions, and forecasting.

We develop a state space model for time series of continuous proportions. Conditionally on the unobserved state, the observations are assumed to follow the Dirichlet distribution, often considered to be the most natural distribution on the simplex. the state follows the Dirichlet-conjugate (DC) distribution which is introduced here. thus the model, while based on the Dirichlet distribution, does not have its restrictive independence properties. the state transition distribution, or "state equation", allows for the incorporation of covariates, trends, seasonality and interventions in a natural way. The model appears to work well when

applied to the tax example.

Consistency of Akaike's Information Criterion for Infinite Variance Autoregressions
Knight, K. (1989)

Suppose X_n is a p-th order autoregressive process with innovations in the domain of attraction of a stable law and the true order p unknown. the estimate of p, $\hat{p}$, is chosen to minimize Akaike's Information Criterion over the integers $0, 1, \ldots, K$. It is shown that $\hat{p}$ is weakly consistent and the consistency is retained if $K \to \infty$ as $N \to \infty$ at a certain rate depending on the index of the stable law.

On the Bootstrap of the Sample Mean in the Infinite Variance Case
Knight, K. (1989)

Athreya (1987) showed that the bootstrap distribution of a sum of infinite variance random variables did not (with probability one) tend weakly to a fixed distribution but instead tended in distribution to a random distribution. In the paper, we give a different proof of Athreya's result motivated by a heuristic large sample representation of the bootstrap distribution.

A Note on the Asymptotic Covariance Matrix of the Yule-Walker Estimator
Knight, K. (1989)

Yohai and Maronna (1977) provided a general form for the asymptotic covariance matrix of the Yule-Walker estimator of autoregressive parameters when the innovations ε_t have a symmetric distribution with $E(\ln^+ |\varepsilon_t|) < \infty$. In this paper, this expression is evaluated for the case where ε_t are in the domain of attraction of a stable law.

Simple Robust Fixed Lag Smoothing
Le, N.D. and Martin, R.D. (1988)

This paper introduces a class of robust lag-k smoothers based on simple low order Markov models for the Gaussian trend-like component of signal plus non-Gaussian noise models. The pth order Markov models are of the pth difference form $\Delta^p x_t = \varepsilon_t$ where $\Delta x_t = x_t - x_{t-1}$ and ε_t is a zero-mean white Gaussian noise process with variance σ_ε^2. The nominal additive noise is a zero-mean white Gaussian noise sequence with variance σ_o^2, while the actual additive noise is non-Gaussian with an outliers generating distribution, e.g., $(1-\gamma)N(0,\sigma_o^2) + \gamma H$. This setup is particularly appropriate for radar glint noise. Implementation of the smoothers requires estimation of the two parameters σ_ε^2 and σ_o^2. This is accomplished using a robustified maximum likelihood approach. Application to both artificial data sets and to glint noise data reveals that the approach is quite effective. We briefly discuss the choices of lag k for the smoothers and also briefly study the sensitivity of our approach to model mismatch.

Modeling Outliers, Burst, and Flat Stretches in Time Series Using Mixture Transition Distribution (MTD) Models
Le, N.D., Martin, R.D., and Raftery, A.E. (1990)

The class of *Mixture Transition Distribution* (MTD) time series models is introduced. In these models, the conditional distribution of the current observation given the past is a mixture of conditional distributions given each one of the last p observations. They can capture non-Gaussian and non-linear features such as outliers, bursts of activity and flat stretches, in a single unified model class. They can also represent time series defined on arbitrary state spaces, which need not even be Euclidean. They perform well in the usual case of Gaussian time series without obvious non-standard behaviors. The models are simple, analytically tractable, easy to simulate and readily estimated.

The stationarity and autocorrelation properties of the models are derived. A simple EM algorithm is given and shown to work well for estimation. The models are applied to several real and simulated data sets with satisfactory results. They appear to capture the features of the data better than the best competing ARIMA models.

Min-max Variance of M-Estimates When Scale is Unknown
Li, B. and Zamar, R.H. (1989)

This paper extends Huber's (1964) min-max result to the case when the scale parameter is unknown and must be estimated along with the location parameter. A min-max problem in which nature chooses F from a family $\mathbf{F}$ of symmetric distribution functions around a given location-scale central model, and the statistician chooses an M-estimate of location, that is, specifies the influence curve or score function ψ and the auxiliary scale estimate s_n, is solved.

The optimal choice for s_n is an M-estimate of scale applied to the residuals about the median. The optimal choice for the score function ψ is a truncated and rescaled maximum likelihood score function for the central model. In the Gaussian case rescaling is not necessary and so, except for the truncation point which is now smaller, Huber's (1964) classical result obtains.

A New Class of Bias Robust Estimates of Regression
Maronna, R.A. and Yohai, V.J.(1989)

In this paper we present a new class of bias-robust estimates of multivariate regression. These estimates are based on a robust estimate of univariate regression through the origin. If y and x are two real random variables, let $T(y, x)$ be the univariate estimate of the regression of y on x. The regression estimate $\mathbf{T}(y, \mathbf{x})$ of a random variable y on a random vector $\mathbf{x} = (x_1, \ldots, x_p)'$ is defined as the value $\mathbf{t} \in R^p$ which minimizes $\sup_{\|\lambda\|} |T(y - \mathbf{t}'\mathbf{x}, \lambda'\mathbf{x})| s(\lambda'\mathbf{x})$ where s is a robust estimate of scale. These estimates, which are called projection estimates, are regression, affine and scale equivariant. When the univariate regression estimate is $T(y, x) = \text{median}(y/x)$, the resulting projection estimate is highly bias-robust. In fact, we found an upper bound for its maximum bias in a contamination neighborhood, which is approximately twice the minimum possible value of this

maximum bias for any regression and a fine equivariant estimate. The maximum bias of this estimate in a contamination neighborhood, compares favorably with Rousseeuw's least median squares estimate and the most bias-robust GM-estimate. We also give a modification of this projection estimate which is almost bias-minimax in the class of all regression and affine equivariant estimates. We prove that the projection estimates have a rate of convergence on $n^{-1/2}$ and give an huristic derivation of their asymptotic distribution, which is non-normal.

Min-max Bias Robust Regression
Martin, R.D., Yohai, V.J., and Zamar, R.H. (1987).

This paper considers the problem of minimizing the maximum asymptotic bias of regression estimates over ε-contamination neighborhoods for the joint distribution of the response and carriers. Two classes of estimates are treated: (i) M-estimates with bounded function ρ applied to the scaled residuals, using a very general class of scale estimates, and (ii) bounded influence function type generalized M-estimates. Estimates in this first class are obtained as the solution of minimization problem, while estimates in the second class are specified by an estimating equation. The first class of M-estimates is sufficiently general to include both Huber Proposal 2 simultaneous estimates of regression coefficients and residuals scale, and Rousseeuw-Yohai S-estimates of regression. It is shown that an S-estimate based on a jump-function type ρ solves the min-max bias problem for the minimization of the α-quantile of the squared residuals, where $\alpha = \alpha(\varepsilon)$ depends on the fraction of contamination ε. When $\varepsilon \to 0.5$, $\alpha(\varepsilon) \to 0.5$ and the min-max estimator introduced by Rousseeuw. For the bounded influence class of GM-estimates, it is shown the "sign" type nonlinearity yields the min-max estimate. This estimate coincides with the minimum gross-error sensitivity GM-estimate. For $\rho=1$, the optimal GM-estimate is optimal among the class of all equivariant regression estimates. The min-max S-estimator has a breakdown point which is independent of the number of carriers ρ and tends to 0.5 as ε increases to 0.5, but has a slow rate of convergence. The min-max decreases to zero with increasing ρ. finally, we compare the min-max biases for both types of estimates, for the case where the nominal model is multivariate normal.

Asymptotically Min-max Bias Robust M-estimates of Scale for Positive Random Variables
Martin R.D., and Zamar, R.H. (1987).

Asymptotically min-max bias robust M-estimates of scale are obtained for positive random variables with ε-contaminated distributions, using any one of a broad class of loss functions of positive and negative bias. Any such estimate is a scaled order statistic, with the order statistic and scaling constant determined by ε, the nominal distribution, and the loss function. Some calculations for specific cases show that for a large range of ε the min-max order statistic solution has breakdown point near .5 and is well approximated by a saled median. Some calculations and Monte Carlo results indicate that the asymptotic result has considerable finite sample-size relevance by virtue of the squared bias being at least as large as the variance for rather modest sample sizes (depending on ε) and by virtue of attractive

mean squared error performance in the presence of even rather small fractions of contamination.

Bias Robust Estimation of Scale with Nuisance Location Parameter
Martin, R.D. and Zamar, R.H. (1988)

Bias Robust Estimation of Scale Where Location is Unknown
Martin, R.D. and Zamar, R.H.(1989)

In this paper we consider the problem of robust estimation of the scale of the location residuals when the "true" underlying distribution of the data belongs to a contamination neighborhood of a parametric location-scale family.

First we show that a scaled version of the MADAM (median of absolute residuals about the median) is approximately most bias-robust within the class of Huber's proposal II joint estimates of location and scale. Then we consider the larger class of M-estimates of scale with general location and show that a scaled version of the SHORTH (the shortest half of the data) is approximately most bias-robust in this case. The exact min-max asymptotic bias estimate is a scaled order statistic of the residuals about a certain location estimate. The exact order, scaling and location depend on the fraction of contamination, the loss function and the central parametric model.

Finally, we present the results of a Monte Carlo simulation study showing that the scaled SHORTH has attractive finite sample mean square error properties for contaminated Gaussian data.

A Continuous Multivariate Exponential Distribution that is Multivariate Phase Type.
O'Cinneide, C.A. and Raftery, A.E. (1989).

A reliability model for a bivariate exponential distribution is extended to the multivariate case. The multivariate exponential distribution is then shown to be of multivariate phase type (MPH), and its MPH representation is derived.

Analysis of a Simple Debugging Model
Raftery, A.E. (1988).

A system has an unknown number of faults. Each fault causes a failure of the system, and is then located and removed. The failure times are independent exponential random variables with common mean. A Bayesian analysis of this model is presented, with emphasis on the situation where vague prior knowledge is represented by limiting, improper, prior forms. This provides a test for reliability growth, estimates of the number of faults, an evaluation of current system reliability, a prediction of the time to full debugging, and a model checking procedure. Three examples are given.

Inference for the Binomial N Parameter: A Hierarchical Bayes Approach
Raftery, A.E. (1988).

A hierarchical Bayes approach to the problem of estimating N in the binomial distribution is presented. This provides a simple and flexible way of specifying prior information, and also allows a convenient representation of vague prior knowledge. It yields solutions to the problems of interval estimation, prediction and decision making, as well as that of point estimation. The Bayes estimator compares favorably with the best, previously proposed, point estimators in the literature.

Are Ozone Exceedance Rates Decreasing?
Raftery, A.E. (1989)

R.L. Smith (1989) in his *Statistical Science* discussion paper, proposed new methods for analyzing extreme values based on the point process view of high-level exceedances, and illustrated them with a detailed analysis of ozone data from Houston, Texas. The methods are powerful and, in particular, the point process of cluster peaks over a high threshold provides a remarkable condensation of the massive data set that he analyses.

Smith's conclusion is that there is no trend in the overall levels of the series, but that there is a marked downward trend in the extreme values. It seems hard to find physical explanations for this, and here the evidence is reassessed in terms of a comparison between competing models for the intensity of a Poisson process. This suggests that there is some evidence for a decreasing trend in exceedance rates but that it is rather weak. If there is a trend, it seems more likely to consist of a fairly abrupt change than a gradual decrease. The possibility of long-memory dependence is also considered and clustering method used is discussed.

Space-time Modeling with Long-memory Dependence: Assessing Ireland's Wind Power Resource
Raftery, A.E. and Haslett, J. (1989)

We consider estimation of the long term average power output from a wind turbine generator at a site for which few data on wind speeds are available. Long term records of wind speeds at the 12 synoptic meteorological stations are also used. Inference is based on a simple and parsimonious approximating model which accounts for the main features of wind speeds in Ireland, namely seasonal effects, spatial correlation, short-memory temporal auto-correlation and long-memory temporal dependence. It synthesizes deseasonalization, kriging, ARMA modeling and fractional differencing in a natural way. A simple kriging estimator performs well as a point estimator, and good interval estimators result from the model. The resulting procedure is easy to apply in practice.

Bounded-influence Estimation in the Errors-in Variables Setup
Yohai, V.J. and Zamar, R.H. (1989)

The problem of robust and bounded influence estimation in the measurment error model is considered. In particular, the nonuniqueness of the solution of completely bounded estimating equations is pointed out, and some possible remedies, which include the use of

reweighted orthogonal regression and one-step Newton-Raphson estimates, are discussed.

Bounded Influence Estimation in the Errors-in-Variables Model.
Zamar, R.H. (1989).

In this paper we introduce the orthogonal regression (OR) analogue of M-estimates of regression which we call orthogonal regression M-estimates (ORM). We show that these estimates are consistent at elliptical error-in-variables models and that they are robust provided the corresponding loss function is bounded. We also consider here the OR analogue of S-estimates of regression which we call orthogonal regression S-estimates (ORS). An important by-product of ORS-estimates is a very robust estimate of the scale of the orthogonal residuals which is crucial for the computation of ORM-estimates. Finally we present an algorithm to compute ORS and ORM estimates and the results of a small Monte Carlo simulation.

5. PUBLICATIONS AND REPORTS

Papers published in refereed journals:

Bruce, A.G., and Martin, R.D. (1989). "Leave-k-out diagnostics for time series (with discussion)." *J. Roy. Statist. Soc.* B, 363-424.
[Note: This is a Royal Statistical Society Discussion Paper]

Knight, K. (1989). "Consistency of Akaike's information criterion for infinite variance autoregressions." *Annals of Statistics*, 17, pp.824-840.

Knight, K. (1989). "On the bootstrap of the sample mean in the infinite variance case." *Annals of Statistics*, 17, pp.1168-1175.

Knight, K. (1989). "A note on the asymptotic covariance matrix of the Yule-Walker estimator." *Statistics and Probability Letters*, 8, pp.407-410.

Martin, R.D., Yohai, V.J., and Zamar, R.H. (1987). "Min-max bias robust regression." *Annals of Statistics*, 17, No. 4, 1608-1630.

Martin R.D., and Zamar, R.H. (1987). "Asymptotically min-max bias robust M-estimates of scale for positive random variables." *Jour. Amer. Stat. Assoc.*, 84, 494-501.

O'Cinneide, C.A. and Raftery, A.E. (1989). "A continuous multivariate exponential distribution that is multivariate phase type."
Statistics and Probability Letters, 7, 323-325.

Raftery, A.E. (1988). "Analysis of a Simple Debugging Model." *Jour. Royal Statist. Soc.*, series C (Applied Statistics), 37, 12-22.

Raftery, A.E. (1988). "Inference for the binomial N parameter: A hierarchical Bayes approach," *Biometrika*, 75, 223-228.

Raftery, A.E. (1989). "Are ozone exceedance rates decreasing?" *Statistical Science.*, 4, 378-381.

Raftery, A.E. and Haslett, J. (1989). "Space-time modeling with long-memory dependence: Assessing Ireland's wind power resource (With Discussion)" *Journal of the Royal Statistical Society*, series C (Applied Statistics), 38, 1-50.
[Note: This is a Royal Statistical Society Discussion paper.]

Zamar, R.H. (1989). "Robust estimation in the errors in variables model." *Biometrika.* 76, 149-60.

Papers Submitted to Refereed Journals (and not yet published):

Banfield, J.D. and Raftery, A.E. (1989). "Ice floe identification in satellite imagery using mathematical morphology and clustering about principal curves". Revised and resubmitted to *Jour. Amer. Statist. Assoc.*

Banfield, J.D. and Raftery, A.E. (1989). "Model-based Gaussian and non-Gaussian clustering" Under revision for *Biometrics*.

Cox, D. (1987). "Gaussian likelihood estimation for nearly nonstationary AR(1) processes." To appear in *Annals of Statistics*, 1989.

Cox, D. (1988). "An analysis of Bayesian inference for nonparametric regression." To appear in *Annals of Statistics*.

Cox, D. and Llatas, E. (1988). "M-estimates for nearly nonstationary autoregression." To appear in *Annals of Statistics*.

Grunwald, G.K., Raftery, A.E., and Guttorp, P. (1989). "Ergodicity and recurrence for exponential family state space models". Submitted to *Journal of Applied Probability*.

Grunwald, G.K., Guttorp, P. and Raftery, A.E. (1989). "Time Series of continuous proportions." Revised and resubmitted to *Journal of the Royal Statistical Society, series B*.

Le, N.D., Martin, R.D., and Raftery, A.E. (1990). "Modeling outliers, bursts, and flat stretches in time series using mixture transition distribution (MTD) models." Submitted to *J. Roy. Statist. Soc., series B*

Martin, R.D., and Zamar, R.H. (1988). "Bias robust estimation of scale with nuisance location parameter." Submitted to *Jour. Amer. Statist. Assoc.*

Technical reports published or non-refereed journals:

Banfield, J.D. and Raftery, A.E. *(1989)*. *"Ice floe identification in satellite imagery by clustering about principal curves and mathematical morphology"*. Tech. Rep. No. 172, Dept. of Statistics, Univ. of Washington, Seattle, WA.

Banfield, J.D. and Raftery, A.E. (1989). "Model based Gaussian and non-Gaussian clustering." Tech. Rep. No. 186, Dept. of Statistics, Univ. of Washington, Seattle, WA.

Cox, D. and Martin, R.D. (1988). "Estimation of score functions: Bikernel estimates and cross-validation." To be submitted to *Jour. Amer. Stat. Assoc.*,
[Also partially supported by ONR contract N00014-86-K-0819.]

Fraiman, R., and Martin, R.D. (1988). "Approximate non-Gaussian conditional mean filters." To be submitted to *IEEE Trans. on Auto. Control*.

Grunwald, G.K., Guttorp, P. and Raftery, A.E. (1989). "Recurrence and ergodicity for exponetial family state space models". Tech. Rept. No. 173, Dept. of Statistics, Univ. of Washington, Seattle, WA.

Le, N.D. and Martin, R.D. (1988). "Simple robust fixed log smoothing." Tech. Rep. No. 149, Dept. of Statistics, Univ. of Washington, Seattle, WA.
[also partially supported by ONR contract N00014-86-K-0819]

Li, B. and Zamar, R.H. (1989). "Min-max variance of M-estimates when scale is unknown," Tech. Rept. No. 75, Dept. of Statistics, Univ. of British Columbia.

Maronna, R.A. and Yohai, V.J. (1989). "A new class of bias robust estimates of regression." Tech. Rept. No. 177, Dept. of Statistics, Univ. of Washington, Seattle, WA.

Martin, R.D. and Zamar, R.H. (1989). "Bias robust estimation of scale where location is unknown." Tech. Rept. No. 184, Dept. of Statistics, Univ. of Washington, Seattle, WA.

Yohai, V.J. and Zamar, R.H. (1989). "Bounded influence estimation in the errors-in-variables model." Proceedings of Arcata Conference on Measurement Errors.

6. **GRADUATE STUDENTS SUPPORTED**

Burns, Patrick	Ph.D., December 1988
Huang, Jian	Ph.D., expected 1991
Le, Nhu Dinh	Ph.D., April 1990
Qin, Yongfeng	Ph.D., expected 1992
Taplin, Ross	Ph.D., December 1990